dbGaP Study Release Notes

Release Notes for NHLBI TOPMed - NHGRI CCDG VAFAR, phs000997.v5.p2
"NHLBI TOPMed - NHGRI CCDG: The Vanderbilt AF Ablation Registry"

For any questions or comments, please contact: dbgap-help@ncbi.nlm.nih.gov.

October	17, 2016	Version 1 Data set release date
September	18, 2017	Version 2 Data set release date
May	29, 2018	Version 3 Data set release date
March	6, 2020	Version 4 Data set release date
May	27, 2021	Version 5 Data set release date

2021-05-27
Version 5 Data set release for NHLBI TOPMed - NHGRI CCDG VAFAR now available

This release includes the addition of Freeze 9 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), and VCFs derived from WGS. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

Data Type	subjects	samples
Phenotype	173	173
Seq_DNA_SNP_CNV (VCFs)	173	173
WGS*	173	173

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

   NHLBI TOPMed WGS VAFAR version 4 phs000997.v4.p2 has been updated to version 5. The dbGaP accession for the current set of data is **phs000997.v5.p2**. The participant number (p#) has not changed in version 5. No new subjects have been added to this study.

2. There are no updates to the phenotype datasets.

Molecular Data Updates

Two genomic accessions, phg001350.v1 freeze 8 and phg001563.v1 freeze9, are associated with the study.
1. See download components 'sample-info' for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in 'genotype-qc' tars.

phg	freeze	sample_cnt	subject_cnt
phg001350.v1	8	173	173

phg001563.v1	9	172	172

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000997/phs000997.v5.p2

2020-03-06

Version 4 Data set release for NHLBI TOPMed - NHGRI CCDG VAFAR now available

This release includes the addition of Freeze 8 whole genome sequences (WGS) brokered through the Sequence Read Archive (SRA), VCFs derived from WGS. TOPMed and CCDG VAFAR have been combined into a single dbGaP study. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

Data Type	subjects	samples
Phenotype	173	173
Seq_DNA_SNP_CNV (VCFs)	173	173
WGS*	173	173

*These data are brokered through the Sequence Read Archive (SRA). Please see Authorized Access instructions below.
For a description of non-SRA SAMPLE_USE terms, please see:
https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi

Study and Phenotype Data Updates

1. **New Study Accession**

   NHLBI TOPMed WGS VAFAR version 3 phs000997.v3.p2 has been updated to NHLBI TOPMed - NHGRI CCDG VAFAR version 4. The dbGaP accession for the current set of data is **phs000997.v4.p2**. The participant number (p#) has not changed in version 4. No new subjects have been added to this study.

2. **Updated Datasets (n=1 dataset)**

pht	version	Dataset Name

5089	4	TOPMed_WGS_VAFAR_Sample_Attributes

3. **Updated Variables (n=4 variables)**

pht	pht version	phv	phv version	Variable Name
5089	4	253610	4	BODY_SITE
5089	4	310118	3	SEQUENCING_CENTER
5089	4	310119	3	Funding_Source
5089	4	310120	3	TOPMed_Phase

Molecular Data Updates

1. See download components 'sample-info' for manifest of genotyped samples and files.
2. Genotype calls are available in matrix format ('genotype-calls-vcf') as VCF file(s) with samples of same data sharing consent.
3. Quality control data are in 'genotype-qc' tars.
4. Only Freeze 5b and Freeze 8 VCFs will be available for download.

phg	freeze	sample_cnt	subject_cnt
phg001061.v1	5b	154	154
phg001350.v1	8	173	173

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000997/phs000997.v4.p2

2018-05-29

Version 3 Data set release for NHLBI TOPMed WGS VAFAR now available

This release includes a second genotype call set (GRCh38) and updated phenotype data. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

	Phenotype	Seq_DNA_SNP_CNV (VCFs)	Seq_DNA_WholeGenome
subjects	173	163	118 and ongoing

samples	173	163	118 and ongoing

Molecular data descriptions:
(https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi)

a. Seq_DNA_WholeGenome: Whole genome sequencing
b. Seq_DNA_SNP_CNV: SNP and CNV genotypes derived from sequence data (VCFs)

Study and Phenotype Data Updates

1. **New Study Accession**

   NHLBI TOPMed WGS VAFAR version 2 phs000997.v2.p2 has been updated to Version 3. The dbGaP accession for the current set of data is **phs000997.v3.p2**. The participant number (p#) has not changed in version 3. New subjects have been added to this study.

2. **Updated Datasets (n=4; all existing variables are updated)**

pht	version	Dataset Name
5087	3	TOPMed_WGS_VAFAR_Subject
5088	3	TOPMed_WGS_VAFAR_Sample
5089	3	TOPMed_WGS_VAFAR_Sample_Attributes
5688	3	TOPMed_WGS_VAFAR_Subject_Phenotypes

3. Please note we are discontinuing the submission and distribution of the SAMPLE_USE variable. The sample use counts will be populated by SRA (sequences) and dbGaP (all other submitted molecular data).

Molecular Data Updates

1. Genotype data are accessioned under phg001061.v1. Please see “sample-info” component for genotyped samples, consent status and mapping of sample-vcf file.
2. Genotypes from whole genome sequencing of 154 samples are available in originally submitted variant-call-format (VCFv4.2). They are mapped to reference genome build GRCh38 and are packed into a folder marked as “genotype-calls-vcf”.
3. QC results, including both submitter provided quality data and consistency-checking results from dbGaP, are in the folder “phg001061.v1.TOPMed_WGS_VAFAR_v3.genotype-qc.WGS_markerset_grc38.MULTI.tar.gz”.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted

participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000997/phs000997.v3.p2

2017-09-18

Version 2 Data set release for NHLBI TOPMed WGS VAFAR now available

This release includes updated phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Additionally, phenotype tables include subjects and samples beyond TOPMed Phase I in order to instantiate IDs for future versions. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

	Phenotype	Seq_DNA_SNP_CNV (VCFs)	Seq_DNA_WholeGenome
subjects	171	117	118
samples	171	117	118

Molecular data descriptions:
(https://www.ncbi.nlm.nih.gov/projects/gap/submission/GetSampleUseTypes.cgi)

a. Seq_DNA_WholeGenome: Whole genome sequencing
b. Seq_DNA_SNP_CNV: SNP and CNV genotypes derived from sequence data (VCFs)

Study and Phenotype Data Updates

1. **New Study Accession**

NHLBI TOPMed WGS VAFAR version 1 phs000997.v1.p1 has been updated to Version 2. The dbGaP accession for the current set of data is **phs000997.v2.p2**. The participant number (p#) has changed in version 2; subjects have been retired. New subjects have been added to this study.

2. **Updated Datasets (n=4; all existing variables are updated)**

pht	version	Dataset Name
5087	2	TOPMed_WGS_VAFAR_Subject
5088	2	TOPMed_WGS_VAFAR_Sample
5089	2	TOPMed_WGS_VAFAR_Sample_Attributes
5688	2	TOPMed_WGS_VAFAR_Subject_Phenotypes

3. **New Variables (n=5)**

pht	pht version	Dataset Name	phv	Variable Name
5089	2	TOPMed_WGS_VAFAR_Sample_Attributes	310118	SEQUENCING_CENTER

5089	2	TOPMed_WGS_VAFAR_Sample_Attributes	310119	Funding_Source
5089	2	TOPMed_WGS_VAFAR_Sample_Attributes	310120	TOPMed_Phase
5089	2	TOPMed_WGS_VAFAR_Sample_Attributes	310121	TOPMed_Project
5089	2	TOPMed_WGS_VAFAR_Sample_Attributes	310122	Study_Name

Molecular Data Updates

1. Genotype data are accessioned under phg000931.v1. Please see “sample-info” component for genotyped samples, sample consent status, and mapping of sample-vcf files.
2. Genotypes from whole genome sequencing are available in originally submitted variant-call-format (VCFv4.2), which are packed in the folder marked as “genotype-calls-vcf”.
3. QC results, including both submitter provided quality data and consistency-checking results of dbGaP, are in the folder “phg000931.v1.TOPMed_WGS_VAFAR_v2.genotype-qc.WGS_markerset_grc37.MULTI.tar.gz”.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and Sequence Read Archive (SRA) data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000997/phs000997.v2.p2

2016-10-17

Version 1 Data set release for NHLBI TOPMed WGS VAFAR now available

This release includes TOPMed Phase I phenotype tables, whole genome sequences (WGS) brokered through the SRA, and VCFs derived from WGS. Additionally, phenotype tables include subjects and samples beyond TOPMed Phase I in order to instantiate IDs for future versions. Please refer to the latest study configuration report for a detailed description of each download component.

Consent group 1 (c1): Health/Medical/Biomedical (IRB) (HMB-IRB)

	phenotype	SRA/VCFs
subjects	122	55
samples	122	55

Molecular Data Updates

1. Genotype data are accessioned under phg000797.v1. Please see “sample-info” component for genotyped samples, sample consent status and mapping of sample-vcf files.
2. Genotypes from whole genome sequencing are available in originally submitted variant-call-format (VCFv4.2), which are packed in the folder marked as “genotype-calls-vcf”.
3. QC results, including both submitter provided quality data and consistency-checking results of dbGaP, are in the folder “phg000797.v1.TOPMed_WGS_VAFAR.genotype-qc.WGS_markerset_grc37.MULTI.tar.gz”.

Authorized Access (Individual Level Data and SRA Data)

Individual level data and SRA sequencing data are available for download through the dbGaP Authorized Access System upon approval of the Data Access Request (DAR):

- https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?login=&page=login

Public FTP site (Summary Level Data Only)

All data tables, data dictionaries, and documents will be housed under one directory for ease of downloading. The data_dict filenames have an added study version number (phs#.v#) and deleted participant set number (p#) from the table accession (pht#.v#). The var_report filenames have an added study version number (phs#.v#). In the var_report files, variables contain version numbers (phv#.v#) and summaries were created for each consent group (c#). These FTP files are available at:

- https://ftp.ncbi.nlm.nih.gov/dbgap/studies/phs000997/phs000997.v1.p1